CMSC 601
Writing: Process and Style

Adapted from slides by
Prof. Marie desJardins

February 2011

Sources
• Also useful: Lyn Dupré, BUGS in Writing. Addison Wesley, 1995.

Questions
• How many of you like to write [in English]?
• How many of you think you’re good at writing [in English]?
• How many of you are worried about writing [for this class, for your thesis/dissertation]?

Overview
• Process:
  – Organizing yourself and your thoughts
• Writing:
  – General stylistic guidelines
  – Specific (but important) suggestions
The Writing Process

• Writing should be part of the research process
  -- It’s really hard to “Do The Work” and then “Write It Up”
  -- For one thing, The Work is never done, and it is constantly changing
  -- Writing helps to pin down the details, and helps to focus your ongoing research

Publishing Papers

• You should publish papers along the way to getting your degree (definitely true for Ph.D. students; ideally true for M.S. students)
• Peters says (p. 217):
  -- “When deciding whether to use the paper publication strategy, be aware that you may have to put in more total work than if you do not publish.”
• BUT:
  -- In CS, you’re expected to have publications when you graduate
  -- Publications are part of the ongoing department evaluation process
  -- The “extra work” more than repays itself in the long term, by focusing your research, and by helping you learn how to write (and how to do publishable research)

Peer Reviewed Papers

• Peer reviewed papers validate your work
  -- Getting papers on your research accepted at a workshop, conference or journal is evidence that your work is good and makes a contribution
• They provide valuable feedback, even for rejected papers
  -- Referees are experts in your area
  -- Their comments, suggestions and recommendations are free consulting
**Write as You Work**

Writing about papers you read:
- makes writing the related-work part of your dissertation that much easier
- creates a record of your understanding of the paper (because you *will* forget the details)
- helps you to organize and synthesize the threads of the related work
- encourages you to analyze and think about previous work and its limitations

**Procrastination**

- De Quincey, quoted by Peters:
  - “If once a man indulge himself in murder, very soon he comes to think little of robbing; and from robbing he next comes to drinking and Sabbath-breaking, and from that to incivility and procrastination.”
- Procrastination-busters:
  - Write something every day, even if it’s a blog post, scribbles, an outline, a paper summary, or a trivial bit of commentary
  - Reward yourself 😊
  - Write sloppy (sloppily?) and fix it later. (But organize well. Bad organization is much harder to fix later.)

**Thesis Structure**

- Specific structure varies, but in CS you should:
  - Describe the problem
  - Explain why it’s important
  - State how you solved the problem
  - Make explicit claims about your approach
  - Support claims experimentally and/or analytically
  - Place your approach in the context of current and past related work
  - Give directions for future work
- Applies in smaller scale and with variations to proposals and technical papers

**A Minor Quibble**

- Peters suggests (p. 215):
  - “Incidentally, don’t make substantial revisions based on input from only a single committee member, since their instructions will often be contradictory and you should resolve contradictions *before* extensive rewriting.”
- The exception is your advisor! As a general rule, you shouldn’t circulate a draft paper/dissertation to your committee until your advisor has OK’d it.
  - Their reputation is on the line
  - The other committee members shouldn’t have to read a half-baked draft. Your advisor will help you bake it.
A Few End Game Tips

• Be very wary of taking a job before you finish your Ph.D.*
• Get your journal papers into draft form before you leave
• Think about what you want to do next before you go out on interviews.
  – Preferably not “extend my dissertation in six different ways.”

Style
(Writing II)

Robert’s Words of Wisdom

• Keep it brief.
• Break it up.
• Don’t be self-important.
• Start your paragraphs with topic sentences.
• Don’t write a detective novel.
• Don’t try to handle too many ideas at once.
• Use key words.
• Signpost with transitional phrases.
• Repeatedly summarize.
• Avoid passive constructions.
• Avoid adverbs (very).
• Delete double negatives.
• Chop off your first paragraph.
• Read it out loud.
• Read it again cold.
• Move back and forth between word processor and paper.

Quoted from Peters, ch 18, pp. 231—233

Zobel in a Nutshell

Simplicity is key.
Don’t Be Obscure
• Example from Peters (p. 229/230):
  – “[The Environmental Protection Agency] has developed an industry-specific cross-media pollution-abatement model that also estimates the reduction in human health risks attributable to adopting various sets of abatement measures. The model has been applied to the iron and steel industry.”
• Rewrite:
  – “In order to understand how to reduce pollution in some specific industries, the [EPA] has developed a computer model which examines how pollutants in air, water, and other environmental media interact. In addition, the model can estimate how selected measures to reduce pollution would also reduce human health risks. As a trial run, the EPA has used this model to examine pollution reduction in the iron and steel industry.”

Flesch–Kincaid Grade Level Metric

Cut It Out I
The volume of information has been rapidly increasing in the past few decades. While computer technology has played a significant role in encouraging the information growth, the latter has also had a great impact on the evolution of computer technology in processing data throughout the years. Historically, many different kinds of databases have been developed to handle information, including the early hierarchical and network models, the relational model, as well as the latest object-oriented and deductive databases. However, no matter how much these databases have improved, they still have their deficiencies. Much information is in textual format. This unstructured style of data, in contrast to the old structured record format data, cannot be managed properly by the traditional database models. Furthermore, since so much information is available, storage and indexing are not the only problems. We need to ensure that relevant information can be obtained upon querying the database. (Zobel p. 12)

Cut It Out I
The volume of information has been rapidly increasing in the past few decades. While computer technology has played a significant role in encouraging the information growth, the latter has also had a great impact on the evolution of computer technology in processing data throughout the years. Historically, many different kinds of databases have been developed to handle information, including the early hierarchical and network models, the relational model, as well as the latest object-oriented and deductive databases. However, no matter how much these databases have improved, they still have their deficiencies. Much information is in textual format. This unstructured style of data, in contrast to the old structured record format data, cannot be managed properly by the traditional database models. Furthermore, since so much information is available, storage and indexing are not the only problems. We need to ensure that relevant information can be obtained upon querying the database. (Zobel p. 12)
Much information is textual. This unstructured data cannot be managed properly by traditional database models. Furthermore, storage and indexing are not the only problems. We need to ensure that relevant information can be obtained upon querying. (Zobel p. 12)

Can you do better?

As part of their work, they showed that the problem of finding the best total order in a set of given items belongs to the class of NP-hard problems. To be able to find an approximation for the global order, in the paper, they provide a simple greedy algorithm and a second slightly modified algorithm that takes advantage of strongly connected graphs to return an approximation to the best ordering.

(from a student’s paper summary of Cohen et al.)
The contribution of this paper was to describe a method in which a collection of objects can be ordered, using preference judgments. There are two stages in which this ordering is done. First, one creates a “binary preference function” in order to determine how to rank the objects. Then, one uses this function to order the objects.

(from a student’s paper summary of Cohen et al.)

The contribution of this paper is to describe a two-stage method in which a two-stage method in which a for ordering a collection of objects can be ordered, using preference judgments. There are two stages in which this ordering is done. First, one creates in this approach, a “binary preference function” is learned from training data; in order to determine how to rank the objects. Then, one uses this function is used to order the a new set of objects.

(from a student’s paper summary of Cohen et al.)

Don’t write overly long papers, sections, paragraphs, sentences, or words.
Know what each section, paragraph, and sentence is about, and stick to the subject.
Define your terms, and use boldface or another convention to make them stand out.
Expand your acronyms (and use as few as possible).
Explain your math in English.

“The reader is always (well, at least sometimes) right (or at least kinda).”
Don’t get defensive and start making excuses:
– “It’s in there!” [Then why didn’t they notice it?]
– “I didn’t have room!” [Then maybe you should rethink your priorities.]
– “It’s not important!” [But this reader thinks it is. So the paper has to explain it, or convince her that it’s not important.]
But ignore “There’s no future work” comments...*

*Anecdote alert!
Revisiting Paraphrasing

• Be careful about paraphrasing: Zobel p. 25-26:
  – According to Fier and Byke such an approach is “simple and...fast, [but] fairly crude and... could be improved”

is revised to
  – Fier and Byke describe the approach as simple and fast, but fairly crude and open to improvement.

• Why skirt the edge of plagiarism when you can use your own words and clarify what is meant:
  – According to Fier and Byke, this approach is efficient, but the quality of the results could be improved.”

On Self-Plagiarism

• OK under certain circumstances...
  – Most researchers reuse parts of earlier papers, especially related work and terminology
  – New publications should be substantially different and/or have significant new results

• Workshop => conference => journal
  – It’s common for a paper to evolve and mature as it is accepted in different kinds of venues
  – There can be a lot of overlap, but there should be new material and/or results at each step

On Style Manuals

• There is no single authority on style. How could there be?
  – E.g., How should we write dates? When should we use words for numbers?

• Most commercial publishers choose an established style manual perhaps with a small collection of “house rules”

• The goal is consistency and to have a way of resolving questions

• Journals that edit papers may enforce their own style rules

Specifics
Example: CMS on Numbers

* Among the factors governing the choice between spelling out numbers and using numerals are whether the number is large or small, whether it is an approximation or an exact quantity, what kind of entity it stands for, and what context it appears in (CMS 2003, 380). Precise measurements are always presented as numerals.

* In nontechnical contexts, the following are spelled out: whole numbers from one through one hundred, round numbers, and any number beginning a sentence. For other numbers, numerals are used (CMS 2003, 380).

* Round numbers. By virtue of their rounding, these numbers are approximate. They are written out. For example, write “The federal deficit was increased by two hundred billion dollars,” or “San Francisco is about twelve hundred miles from Denver.” But also write, “The race followed a straight course from Denver to San Francisco, a distance of 1,255.6 miles.”

* Beginning a sentence. When numbers or a date are required to open a sentence, write them out. For example: “One hundred five girls and sixteen boys tried out for the varsity soccer team.” If you can, rewrite the sentence so it does not begin with a number:

  Mixed numbers? Do not mix numerals with written numbers when they refer to similar things. For example, write “Only 10 of the 150 tourists were willing to visit the city after the riot.” Do not write, “Only ten of the 150 tourists…”

* Mixed sets of numbers. Sometimes two sets of numbers are embedded in a single sentence. For clarity, present one set written out, the other as numerals. For example, write “There were eighty-three contestants who dropped out before covering 250 miles.”

* Numbers & units. Generally, do not mix numbers that are spelled out with symbols, write out the term for the symbols as well. For example, write “the temperature was 45°F, or forty-five degrees; 100 or twenty dollars. Chicago style makes an exception for percentages: it is OK to write 45 percent.

* Compound numbers. Hyphenate compound numbers from twenty-one to ninety-nine, compounds with a number as the first element, and the written form of fractions.

* Ordinal numbers. Follow the general rules as for other numbers. For example, write “The window for applications was the third to twenty-third of August.” But use numerals with ordinal numbers above one hundred. For example, write: “Hade Selassie was the 225th Emperor of Ethiopia.”

* Centuries. Write out references to centuries, the eighteenth century, the twenty-first century, in lowercased letters.

From CMS Crib Sheet, Dr. Abel Scribe, Fall 2007, http://www.docstyles.com/cmscrib.htm

Precriptivism

- Prescriptive vs. descriptive linguistics
  - How language should be used vs. how it is used

- Precriptivism focuses on choices in spelling, words, grammar, pronunciation and syntax

- Precriptivists often go too far and embrace rules that educated and articulate language users do not always follow

- But remember that your reviewer may be a precriptivist

- Be sensitive to the norms in your community

Avoid Slang and Idioms

- Zobel: “crop up,” “lose track,” “it turned out that,” “play up,” “right out,” “run the gamut,” “teased into”
- Also: “lots,” “a lot,” “write up”
- Summaries: “good job,” “come up with,” “it was odd,” “a difficult read”
- Avoid contractions, which are often considered too informal for science writing
- Not common in American English: “viz.,” “the works,” “hence”

Avoid Qualifiers and Adverbs

- “Very,” “rather,” “simply,” “possibly,” “of course,” “naturally,” “obviously,” “just,” “pretty,” “pretty much,” “more of,” “extremely,” “seriously,” “indeed,” “really”
- Particularly avoid qualifying nonqualifiable words such as “unique,” “intractable,” “optimal,” and “infinite”
- Avoid personalizing your remarks: Minimize the use of “I think,” “I feel,” “I believe,” “It seems”
Avoid Fluff: Zobel p 55.

- adding together → adding
- cancel out → cancel
- during the course of → during
- for the purpose of → for
- in view of the fact → given
- the vast majority → most
- a number of → several
- whether or not → whether
- it can be seen that → it is a fact that
- is something that can → can
- involves → (active verb)
- In the paper, ... →
- They noticed that... →

Parallel Construction

- ☑ “I like to go swimming, riding bicycles, and I read a lot.”
- ☑ “The complexity increased both in time and space.”
- ☑ “The three most important things to remember are:
  1. Write a little every day.
  2. You should proofread everything before showing it to your advisor.
  3. Careful of bad grammar!”

Parallel Construction II

- The key findings are:
  - The algorithm to learn the preference function, based on the Freund and Schapire “Hedge” algorithm.
  - An algorithm to find the ordering... is NP-complete; however, ....
  - A system (composed of these two algorithms) to compile the result sets of various searches often performs better than a domain expert entity.
- The key contributions are:
  - An algorithm to learn the preference function, based on Freund and Schapire’s Hedge algorithm.
  - A proof that finding the optimal ordering is NP-complete, and a greedy algorithm that is guaranteed...
  - A system (composed...) that compiles the result sets of multiple searches, and often performs better than any individual search.

Nonsexist Writing

- To avoid the use of “he,” you can:
  - Use the plural
  - Rewrite to avoid pronouns
  - Name people in examples (with alternating male/female names)
  - OK these days to use “they” for singular nouns
Proper References


Some of My* Personal Nits

• *Its* vs. *it's*
• *Which* vs. *that*
  – "which" qualifies (and is always preceded by a comma); "that" defines. Heuristic: Use "that" by default.
  – http://www.kentlaw.edu/academics/lrw/grinker/LwtaThat_Versus_Which.htm
• Between vs. among
• *Dangling "this" references*
• Affect vs. effect
• Continual vs. continuous
• Optimize vs. improve
• Plurals and apostrophes
• Colons, semicolons, and dashes
  • i.e. / e.g. / etc. / et al.
• Hyphenate compound adjectives, *not* adverbs or nouns!
• Commas!

Remarks From Previous Semesters

• Use consistent tense
  – Generally the present tense
• Punctuation goes *inside* quotation marks
• "Scare quotes" vs *italics* to introduce new terms
  – Only do this the *first time* you use the new term
• Avoid passive voice (usually)
  – Authors and algorithms do things; they don’t just happen
  – I vs. we:
    • Use "I" for dissertations, theses, proposals
    • Use "we" for co-authored paper
  • Convention goes either way on single-authored papers (but I personally prefer "I")
• Mostly, try to avoid personal pronouns unless you need them!

Remarks From Previous Semesters II

• The summary itself should be written in a formal, "scientific paper" style
  – The wrapup (discussion of presentation etc.) can be more informal
• "Via" → "using" or "by"
• Semicolon vs. colon vs. comma
• "As" → "because" or "since"